PATENT ABSTRACTS OF JAPAN

(11)Publication number : 09-114484
(43)Date of publication of application : 02.05.1997

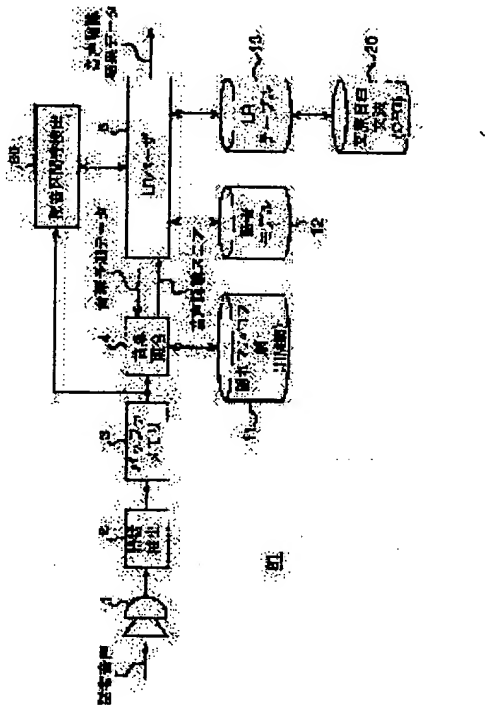
(51)Int.Cl. G10L 3/00
G10L 3/00

(21)Application number : 07-275866
(22)Date of filing : 24.10.1995

(71)Applicant : ATR ONSEI HONYAKU TSUSHIN KENKYUSHO:KK
(72)Inventor : TAKEZAWA TOSHIYUKI
MORIMOTO TAKUMA

(54) VOICE RECOGNITION DEVICE

(57)Abstract:
PROBLEM TO BE SOLVED: To eliminate the ambiguity of modification relation in a syntax analysis by determining the modification relation of the syntax analysis as to each voice section divided with a voiceless section, etc., and recognizing a spoken voice.
SOLUTION: A detection part 30 for a voiceless section, etc., detects the voiceless section, etc., including a section based upon a pause, a redundant word, rhythmical information, etc., according to a time series of feature parameters outputted from a buffer memory 3, and outputs its detection signal to an LR purser 5. The LR purser 5 reads in data in the speech section of a section unit indicated with the inputted detection signal and performs a section-limited HMM-LR process using an HMM-LR method for the speech section. Consequently, the modification relation in the speech section of each section unit is determined, and then modification relation between words, phrases, or clauses in the speech sections of different section units is determined.



LEGAL STATUS

[Date of request for examination]	24.10.1995
[Date of sending the examiner's decision of rejection]	
[Kind of final disposal of application other than the examiner's decision of rejection or application converted registration]	
[Date of final disposal for application]	
[Patent number]	2880436
[Date of registration]	29.01.1999
[Number of appeal against examiner's decision of rejection]	
[Date of requesting appeal against examiner's decision of rejection]	
[Date of extinction of right]	29.01.2003

(19)日本国特許庁（ＪＰ）　(12)**公開特許公報**（Ａ）　(11)特許出願公開番号

特開平9－114484

(43)公開日　平成９年(1997)５月２日

(51)Int.Cl.[6]	識別記号	庁内整理番号	ＦＩ		技術表示箇所
Ｇ１０Ｌ　3/00	５３５		Ｇ１０Ｌ　3/00	５３５	
	５６１			５６１Ｇ	

審査請求　有　　請求項の数４　ＯＬ　（全 13 頁）

(21)出願番号　特願平7－275866

(22)出願日　平成７年(1995)10月24日

(71)出願人　593118597
株式会社エイ・ティ・アール音声翻訳通信研究所
京都府相楽郡精華町大字乾谷小字三平谷５番地

(72)発明者　竹澤　寿幸
京都府相楽郡精華町大字乾谷小字三平谷５番地　株式会社エイ・ティ・アール音声翻訳通信研究所内

(72)発明者　森元　逞
京都府相楽郡精華町大字乾谷小字三平谷５番地　株式会社エイ・ティ・アール音声翻訳通信研究所内

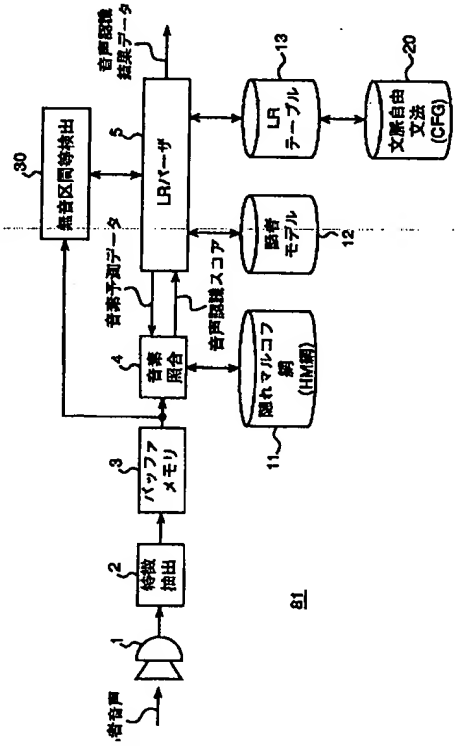
(74)代理人　弁理士　青山　葆　（外２名）

(54)【発明の名称】　音声認識装置

(57)【要約】

【課題】　統語解析における係り受け関係の曖昧性を解消することのできる音声認識装置を提供する。

【解決手段】　入力された発声音声を音声認識して音声認識結果を出力する音声認識手段を備えた音声認識装置において、入力された発声音声に基づいてポーズと冗長語と句又は節の境界とのうちの少なくとも１つを検出して検出信号を出力する検出手段を備え、音声認識手段は、ポーズと冗長語と句又は節の境界とのうちの少なくとも１つによって分割された複数の音声区間からなる入力された発声音声の各音声区間の音声認識処理をした後、異なる音声区間に属する語、句又は節の間の係り受け関係を決定して発声音声の音声認識をする。



1

【特許請求の範囲】

【請求項１】　入力された発声音声を音声認識して音声認識結果を出力する音声認識手段を備えた音声認識装置において、

入力された発声音声に基づいてポーズと冗長語と句又は節の境界とのうちの少なくとも１つを検出して検出信号を出力する検出手段を備え、

上記音声認識手段は、上記検出信号に基づいて統語解析における係り受け関係を決定して上記発声音声の音声認識をすることを特徴とする音声認識装置。

【請求項２】　上記音声認識手段は、上記ポーズと冗長語と句又は節の境界とのうちの少なくとも１つによって分割された複数の音声区間からなる入力された発声音声の各音声区間について音声認識処理をした後、異なる音声区間に属する語、句又は節の間の係り受け関係を決定して、上記入力された発声音声の音声認識をすることを特徴とする請求項１記載の音声認識装置。

【請求項３】　上記検出手段は、上記発声音声のパワーが、所定の時間の範囲だけ、所定のしきい値以下である第１の条件と、上記発声音声のゼロクロスの数が、所定の時間の間において、所定のしきい値以上である第２の条件とのうち少なくとも１つの条件が満足することを検出することにより上記ポーズを検出することを特徴とする請求項１又は２記載の音声認識装置。

【請求項４】　上記検出手段は、上記ポーズと冗長語と句又は節の境界とのうちの少なくとも１つを、それぞれの予め決められた言語モデルに一致するか否かを判断することにより検出することを特徴とする請求項１又は２記載の音声認識装置。

【発明の詳細な説明】

【０００１】

【発明の属する技術分野】本発明は音声認識装置に関し、特に、発声音声中におけるポーズ（無音区間）又は冗長語などの無音区間等を検出して連続的に音声認識を実行する音声認識装置に関する。なお、本明細書では、ポーズと冗長語並びに韻律的な情報等を手がかりとする区切りとを含むものを無音区間等という。

【０００２】

【従来の技術】近年、連続音声認識の研究が盛んに行われ、いくつかの研究機関で文音声認識システムが構築されている。これらのシステムの多くは丁寧に発声された音声を入力対象にしている。しかしながら、人間同士のコミュニケーションでは、「あのー」、「えーと」などに代表される冗長語や、一時的に発声音声が無い無音区間等の状態のポーズである言い淀みや言い誤り及び言い直しなどが頻繁に出現する。

【０００３】図９は、図２に示す例文「きれいな黒い髪の女の子を見た」を従来例の連続音声認識装置で音声認識処理を実行するときの音声認識動作をスタック形式で示す図である。従来例の連続音声認識装置の音声認識動

2

作について図９を参照して説明する。まず、図９の状態スタック２０１に示すように、「きれいな」という発声音声の系列が認識されて文字として積まれる。次に、状態スタック２０１における「きれいな」という文字は音声認識用辞書に載っているので、状態スタック２０２に示すように形容詞句を表す「ａｄｊ」という文字に変換される。次に、「黒い」という発声音声の系列が認識されて状態スタック２０３に示すように文字として積まれ、状態スタック２０３における「黒い」は音声認識辞書に載っているので状態スタック２０４に示すように形容詞句を表す「ａｄｊ」という文字に変換される。

【０００４】次に、「髪の」という発声音声の系列が認識されて状態スタック２０５に示すようにさらに文字として積まれ、状態スタック２０５における「髪の」という文字は音声認識辞書に載っているので状態スタック２０６に示すように名詞句を表す「ＮＰ」という文字に変換される。さらに、状態スタック２０６において、形容詞句の「ａｄｊ」と名詞句の「ＮＰ」とは名詞句の「ＮＰ」になるという構文規則が適用されて、「黒い」が変換された形容詞句の「ａｄｊ」と「髪の」が変換された名詞句の「ＮＰ」とは、状態スタック２０７に示すように名詞句の「ＮＰ」に変換される。すなわち、状態スタック２０７における名詞句の「ＮＰ」は「黒い髪の」を表す。

【０００５】ここで、状態スタック２０７において、形容詞句の「ａｄｊ」と名詞句の「ＮＰ」とは名詞句の「ＮＰ」になるという構文規則を適用するかしないか２つの選択枝がある。ここで、構文規則を適用すると「きれいな」は「髪の」に係ることになり、構文規則を適用しないと「きれいな」は「髪の」に係らない構文構造のままで係り受け関係の決定は以降の音声認識処理に持ち越されることになる。従って、このような場合、従来例の連続音声認識装置では、文字を積む装置を２つに分離して以降の音声認識を実行する。すなわち、一方の装置は、状態スタック２０７に構文規則を適用した状態スタック２０８に示す状態で以降の音声認識処理を実行し、他方の装置は、状態スタック２０７のままの状態で以降の音声認識処理を実行する。ここで、一方の装置の状態スタック２０８における名詞句の「ＮＰ」は「きれいな黒い髪の」を表す。

【０００６】一方の装置において、状態スタック２０９に示すように「きれいな黒い髪の」を表示する名詞句「ＮＰ」の上に、「女の子を」という発声音声の系列が認識されて文字として積まれ、状態スタック２０９における「女の子を」の文字は音声認識辞書に載っているので状態スタック２１０に示すように名詞句を表す「ＮＰ」という文字に変換される。次に状態スタック２１０において、名詞句の「ＮＰ」と名詞句の「ＮＰ」は名詞句の「ＮＰ」になるという構文規則が適用されて、状態

詞句の「ＮＰ」と「女の子を」が変換された名詞句の「ＮＰ」は状態スタック２１１に示すように名詞句の「ＮＰ」に変換される。ここで、状態スタック２１１の名詞句「ＮＰ」は「きれいな黒い髪の女の子」を表す。そして、「見た」という発声音声の系列が認識されて状態スタック２１２に示すように文字として積まれ、状態スタック２１２における「見た」は音声認識用辞書に載っているので状態スタック２１３に示すように動詞句を表す「ＶＰ」に変換され、状態スタック２１４に示すように１つの文章として認識される。すなわち、「きれいな」が「髪に」に係る構造の認識結果が得られる。

【０００７】他方の装置において、「女の子を」という発声音声の系列が認識されて状態スタック２２１に示すように文字として積まれ、「女の子を」の文字は状態スタック２２２に示すように名詞句を表す「ＮＰ」という文字に変換される。次に状態スタック２２２において、構文規則が適用されて、「黒い髪の」が変換された名詞句の「ＮＰ」と「女の子を」が変換された名詞句の「ＮＰ」は状態スタック２２３に示すように名詞句の「ＮＰ」に変換される。ここで、状態スタック２２３の名詞句「ＮＰ」は「黒い髪の女の子」を表す。そして、さらに構文規則が適用されて、「きれいな」が変換された形容詞句の「ａｄｊ」と「黒い髪の女の子を」が変換された名詞句の「ＮＰ」は状態スタック２２４に示すように名詞句の「ＮＰ」に変換される。すなわち、「きれいな」が「女の子を」に係る構造として認識される。次に、「見た」という発声音声の系列が認識されて状態スタック２２５に示すように文字として積まれ、状態スタック２２５における「見た」は状態スタック２２６に示すように動詞句を表す「ＶＰ」に変換され、状態スタック２２７に示すように１つの文章として認識される。すなわち、「きれいな」が「女の子を」に係る構造の認識結果が得られる。

【０００８】

【発明が解決しようとする課題】以上詳述したように、図２の例文を従来例の連続音声認識装置で認識すると、「きれいな」が「髪の」に係る構造の認識結果と、「きれいな」が「女の子を」に係る構造の認識結果の２つの異なる構造の認識結果が得られ、統語解析における係り受け関係の曖昧性が解消できないという問題点があった。また、その結果さらに長い発話を扱うと曖昧性が増していくという問題点があった。

【０００９】本発明の目的は以上の問題点を解決し、統語解析における係り受け関係の曖昧性を解消することのできる音声認識装置を提供することにある。

【００１０】

【課題を解決するための手段】本発明に係る請求項１記載の音声認識装置は、入力された発声音声を音声認識して音声認識結果を出力する音声認識手段を備えた音声認識装置において、入力された発声音声に基づいてポーズ

と冗長語と句又は節の境界とのうちの少なくとも１つを検出して検出信号を出力する検出手段を備え、上記音声認識手段は、上記検出信号に基づいて統語解析における係り受け関係を決定して上記発声音声の音声認識をすることを特徴とする。

【００１１】また、請求項２記載の音声認識装置は、請求項１記載の音声認識装置において、上記音声認識手段は、上記ポーズと冗長語と句又は節の境界とのうちの少なくとも１つによって分割された複数の音声区間からなる入力された発声音声の各音声区間について音声認識処理をした後、異なる音声区間に属する語、句又は節の間の係り受け関係を決定して、上記入力された発声音声の音声認識をすることを特徴とする。

【００１２】さらに、請求項３記載の音声認識装置は、請求項１又は２記載の音声認識装置において、上記検出手段は、上記発声音声のパワーが、所定の時間の範囲だけ、所定のしきい値以下である第１の条件と、上記発声音声のゼロクロスの数が、所定の時間の間において、所定のしきい値以上である第２の条件とのうち少なくとも１つの条件が満足することを検出することにより上記ポーズを検出することを特徴とする。

【００１３】またさらに、請求項４記載の音声認識装置は、請求項１又は２記載の音声認識装置において、上記検出手段は、上記ポーズと冗長語と句又は節の境界とのうちの少なくとも１つを、それぞれの予め決められた言語モデルに一致するか否かを判断することにより検出することを特徴とする。

【００１４】

【発明の実施の形態】以下、図面を参照して本発明に係る実施形態について説明する。

＜第１の実施形態＞図１は、本発明に係る第１の実施形態である連続音声認識装置８１のブロック図である。第１の実施形態の連続音声認識装置８１は、ＳＳＳ（Successive State Splitting：逐次状態分割法）－ＬＲ（left-to-right rightmost derivation型、すなわち最右導出型）不特定話者連続音声認識装置であって、隠れマルコフ網（以下、HM網という。）メモリ１１に格納された隠れマルコフモデル（以下、HMMという。）のネットワークを用いて音素照合処理を音素照合部４で実行しその結果である音声認識スコアを音素コンテキスト依存型ＬＲパーザ（以下、ＬＲパーザという。）５に送り、これに応答してＬＲパーザ５が入力された発声音声の１つの文に対して連続音声認識を実行して音素予測データを音素照合部４に送って音声認識処理を行う。第１の実施形態は特に、バッファメモリ３から出力される特徴パラメータの時系列に基づいてポーズや冗長語並びに韻律的な情報等を手がかりとする区切りを含む無音区間等を検出してその検出信号をＬＲパーザ５に出力する無音区間等検出部３０を備え、ＬＲパーザ５は、無音区間

単位の音声区間のデータを読み込んで、当該音声区間に対してＨＭＭ－ＬＲ法を用いた区間制限付きＨＭＭ－ＬＲ処理を実行し、最後の区切り単位の末端まで到達すると入力された発声音声の１つの文に対して区間制限無しＨＭＭ－ＬＲ処理を実行することにより音声認識結果データを出力することを特徴とする。

【００１５】ここで、上記ＳＳＳにおいては、音素の特徴空間上に割り当てられた確率的定常信号源（状態）の間の確率的な遷移により音声パラメータの時間的な推移を表現した確率モデルに対して、尤度最大化の基準に基づいて個々の状態をコンテキスト方向又は時間方向へ分割するという操作を繰り返すことによって、モデルの精密化を逐次的に実行する。

【００１６】図１において、話者の発声音声はマイクロホン１に入力されて音声信号に変換された後、特徴抽出部２に入力される。特徴抽出部２は、入力された音声信号をＡ／Ｄ変換した後、例えばＬＰＣ分析を実行し、対数パワー、１６次ケプストラム係数、Δ対数パワー及び１６次Δケプストラム係数を含む３４次元の特徴パラメータを抽出する。抽出された特徴パラメータの時系列はバッファメモリ３を介して音素照合部４に入力される。

【００１７】音素照合部４に接続されるＨＭ網メモリ１１内のＨＭ網は、各状態をノードとする複数のネットワークとして表され、各状態はそれぞれ以下の情報を有する。

（ａ）状態番号
（ｂ）受理可能なコンテキストクラス
（ｃ）先行状態、及び後続状態のリスト
（ｄ）出力確率密度分布のパラメータ
（ｅ）自己遷移確率及び後続状態への遷移確率

【００１８】なお、第１の実施形態において、ＨＭ網は、各分布がどの話者に由来するかを特定する必要があるため、所定の話者混合ＨＭ網を変換して作成する。ここで、出力確率密度関数は３４次元の対角共分散行列をもつ混合ガウス分布であり、各分布はある特定の話者のサンプルを用いて学習されている。

【００１９】音素照合部４は、ＬＲパーザ５からの音素照合要求に応じて音素照合処理を実行する。このときに、ＬＲパーザ５からは、音素照合区間及び照合対象音素とその前後の音素から成る音素コンテキスト情報が渡される。音素照合部４は、受け取った音素コンテキスト情報に基づいてそのようなコンテキストを受理することができるＨＭ網上の状態を、先行状態リストと後続状態リストの制約内で連結することによって、１つのモデルが選択される。そして、このモデルを用いて音素照合区間内のデータに対する尤度が計算され、この尤度の値が音素照合スコアとしてＬＲパーザ５に返される。このときに用いられるモデルは、ＨＭＭと等価であるために、尤度の計算には通常のＨＭＭで用いられている前向きパスアルゴリズムをそのまま使用する。

【００２０】一方、無音区間等検出部３０は、バッファメモリ３から出力される特徴パラメータの時系列に基づいてポーズや冗長語並びに韻律的な情報等を手がかりとする区切りを含む無音区間等を検出して、その検出信号をＬＲパーザ５に出力する。ここで、無音区間等検出部３０は、冗長語については予め内部メモリに格納された冗長語の音素モデルと比較照合することにより冗長語として認識する一方、無音区間であるポーズについては以下の２つの条件のうちの１つが満足するときにポーズとして検出する。

（第１の検出条件）パワーが所定のしきい値レベル以下である時間ｔ０が例えば以下の範囲の値のとき。好ましくは、５０ミリ秒≦ｔ０≦３秒。より好ましくは、５０ミリ秒≦ｔ０≦５００ミリ秒。

（第２の検出条件）入力された音声信号がゼロ電位と交差するゼロクロスの数が所定のしきい値以上である時間ｔ１が例えば以下の範囲の値のとき。好ましくは、５０ミリ秒≦ｔ１≦３秒。より好ましくは、５０ミリ秒≦ｔ１≦５００ミリ秒。さらに、韻律的な情報等を手がかりとする区切りとは、具体的には、イントネーションが急激に上昇又は下降するときは、句又は節の境界であると推測される。これについては、入力される特徴パラメータのうち基本周波数が所定の傾斜の度合い以上で急激に上昇し又は下降して変化したことを検出することにより当該区切り又は境界と判別する。

【００２１】そして、ＬＲパーザ５は、無音区間等検出部３０から入力された検出信号で示された区切り単位の音声区間のデータを読み込んで、当該音声区間に対してＨＭＭ－ＬＲ法を用いた区間制限付きＨＭＭ－ＬＲ処理を実行し、最後の区切り単位の末端まで到達すると入力された発声音声の１つの文に対して区間制限無しＨＭＭ－ＬＲ処理を実行することにより音声認識結果データを出力する。ここで、区間制限付きＨＭＭ－ＬＲ処理とは、１つの区切り単位の音声区間内に限って実行するＨＭＭを用いたＬＲパーザ５による音声認識処理のことであり、区間制限無しＨＭＭ－ＬＲ処理とは、区間を限定せず、入力された発声音声の１つの文に対して、異なる区切り単位の音声区間に属する語、句又は節にＬＲテーブルメモリ１３内の構文規則を適用して実行するＨＭＭを用いたＬＲパーザ５による音声認識処理のことである。ここで、音声区間とは図５に示すように入力された発声音声の１つの文のうちの無音区間等（図５においては括弧を付して示している。）によって分割された１つの区間のことをいい、区切り単位とは図５において括弧を付して示すように音声区間と当該音声区間の後にある無音区間等とからなる１単位のことをいう。また、本明細書において、無音区間等とはポーズと冗長語並びに韻律的な情報等を手がかりとする区切りとを含むものをいい、ポーズ単位とは図５に示すようにポーズによって分

7

【００２２】文脈自由文法データベースメモリ２０内の所定の文脈自由文法（ＣＦＧ）は公知の通り予め自動的に変換されてＬＲテーブルを作成してＬＲテーブルメモリ１３に格納される。ＬＲパーザ５は、例えば音素継続時間長モデルを含む話者モデルメモリ１２と上記ＬＲテーブルとを参照して、入力された音素予測データについて左から右方向に、後戻りなしに処理する。構文的にあいまいさがある場合は、スタックを分割してすべての候補の解析が平行して処理される。ＬＲパーザ５は、ＬＲテーブルメモリ１３内のＬＲテーブルから次にくる音素を予測して音素予測データを音素照合部４に出力する。これに応答して、音素照合部４は、その音素に対応するＨＭ網メモリ１１内の情報を参照して照合し、その尤度を音声認識スコアとしてＬＲパーザ５に戻し、順次音素を連接していくことにより、連続音声の認識を行っている。

【００２３】以上のように構成された第１の実施形態の連続音声認識装置８１において、特徴抽出部２と音素照合部４とＬＲパーザ５とは、例えばデジタル電子計算機で構成される。

【００２４】図６は、図１の連続音声認識装置８１のＬＲパーザ５において実行される音声認識処理を示すフローチャートである。以下、図６を参照して音声認識処理について説明する。

【００２５】図６に示すように、ステップＳ１においては、ＨＭＭ作業域の初期化、並びにＬＲパーザ５の初期化を実行する。具体的には、状態スタック０のセルを１個作成する。ここで、連続音声認識装置８１において用いるセルは、従来のＨＭＭ－ＬＲ法の音声認識の解析に必要な情報を保持するデータ構造、すなわち状態スタックを有するＬＲ作業域と、音声認識スコアと確率テーブルとからなるＨＭＭ作業域とを有する。

【００２６】そして、ステップＳ２において、無音区間等検出部３０から入力された検出信号で示された区切り単位の音声区間のデータを読み込む。さらに、ステップＳ３において、音声データが読み込まれた区切り単位の音声区間に対してＨＭＭ－ＬＲ法を用いた区間制限付きＨＭＭ－ＬＲ処理を実行する。ステップＳ４において、複数の区切り単位のうち最後の区切り単位の末端まで到達したか否かが判断され、最後の区切り単位の末端まで到達していないときは（ステップＳ４においてＮＯ）ステップＳ２に進み、ステップＳ２，Ｓ３の処理を繰り返す。一方、ステップＳ４において、最後の区切り単位の末端まで到達しているときは（ステップＳ４においてＹＥＳ）ステップＳ５に進み、区間制限無しＨＭＭ－ＬＲ処理を実行して音声認識処理を終了する。

【００２７】次に、図１の第１の実施形態の連続音声認識装置８１の音声認識動作を図２に示す例文を用いて説明する。図２は、文の構造解析すなわち統語解析における係り受け関係の曖昧性を含む、例文である。図２の例

8

文を文字列のみを認識して解析しようとすると、図２の例文の上に矢印で示した第１の係り受け関係と例文の下に矢印で示した第２の係り受け関係の少なくとも２つの係り受け関係の曖昧性が残る。すなわち、「きれいな」が「女の子」に係る第１の係り受け関係の「きれいな女の子」であるのか、「きれいな」が「髪」に係る第２の係り受け関係の「きれいな髪」であるのかが不明である。本発明者らは、無音区間であるポーズを利用することにより上述の２つの係り受け関係のうちのいずれか１つに決定できることを見いだした。すなわち、「きれいな」と「黒い」との間に無音区間であるポーズ（図２においては、「きれいな」と「黒い」との間に「△」で示している。）があれば、「きれいな」が「女の子」に係る第１の係り受け関係であると決定でき、「髪の」と「女の子を」との間にポーズ（図２においては、「髪の」と「女の子を」との間に「△」で示している。）があれば、「きれいな」が「髪」に係る第２の係り受け関係であると決定できる。本発明は上述のポーズと係り受け関係との間の規則を利用して、統語解析における係り受け関係の曖昧性を取り除いて音声認識処理を実行している。

【００２８】図３は、図２の例文において第１の係り受け関係を有する場合の連続音声認識装置８１の音声認識動作をスタック形式で示す図である。以下に第１の係り受け関係を有する場合の音声認識動作を図３を参照して説明する。まず、図３の状態スタック５１に示すように、ＬＲパーザ５で「きれいな」という発声音声の系列が認識されて文字として積まれ、次に「きれいな」の認識処理の直後でポーズが無音区間等検出部３０によって検出されて、検出信号が当該検出部３０からＬＲパーザ５に入力されて「きれいな」という文字の上にポーズを表示する「△」として積まれる。次に、状態スタック５１における「きれいな」という文字は音声認識用辞書に載っているので、状態スタック５２に示すように形容詞句を表す「ａｄｊ」という文字に変換される。次に、ＬＲパーザ５で「黒い」という発声音声の系列が認識されて状態スタック５３に示すようにポーズを表示する「△」の上に文字として積まれ、状態スタック５３における「黒い」は音声認識辞書に載っているので状態スタック５４に示すように形容詞句を表す「ａｄｊ」という文字に変換される。ここで、状態スタック５４において「きれいな」が変換された形容詞句の「ａｄｊ」と「黒い」が変換された形容詞句の「ａｄｊ」とには、間にポーズを表示する「△」が積まれているので構文規則は適用されない。

【００２９】次に、ＬＲパーザ５で「髪の」という発声音声の系列が認識されて状態スタック５５に示すように「黒い」が変換された形容詞句の「ａｄｊ」の上に文字として積まれ、状態スタック５５における「髪の」という文字は音声認識辞書に載っているので状態スタック

６に示すように名詞句を表す「ＮＰ」という文字に変換される。さらに、状態スタック５６において、形容詞句の「ａｄｊ」と名詞句の「ＮＰ」とは名詞句の「ＮＰ」になるという構文規則が適用されて、「黒い」が変換された形容詞句の「ａｄｊ」と「髪の」が変換された名詞句の「ＮＰ」とは状態スタック５７に示すように名詞句の「ＮＰ」に変換される。すなわち、状態スタック５７における名詞句の「ＮＰ」は「黒い髪の」を表す。次に、「女の子を」という発声音声の系列が認識されて状態スタック５８に示すように「黒い髪の」を表す名詞句の「ＮＰ」の上に文字として積まれ、状態スタック５８における「女の子を」の文字は音声認識辞書に載っているので状態スタック５９に示すように名詞句を表す「ＮＰ」という文字に変換される。

【００３０】次に状態スタック５９において、名詞句の「ＮＰ」と名詞句の「ＮＰ」は名詞句の「ＮＰ」になるという構文規則が適用されて、状態スタック５９の「黒い髪の」が変換された名詞句の「ＮＰ」と「女の子を」が変換された名詞句の「ＮＰ」は状態スタック６０に示すように名詞句の「ＮＰ」に変換される。ここで、状態スタック６０の名詞句の「ＮＰ」は「黒い髪の女の子」を表す。そして、ＬＲパーザ５で「見た」という発声音声の系列が認識されて状態スタック６１に示すように「黒い髪の女の子」を表す名詞句の「ＮＰ」の上に文字として積まれ、状態スタック６１における「見た」は音声認識用辞書に載っているので状態スタック６２に示すように動詞句を表す「ＶＰ」に変換される。

【００３１】そして、最後のポーズ単位の末端まで到達していると判断されて、ポーズを表示する「△」の前後に位置する「きれいな」を表す形容詞句の「ａｄｊ」と「黒い髪の女の子を」を表す名詞句の「ＮＰ」とに、形容詞句の「ａｄｊ」と名詞句の「ＮＰ」とは名詞句の「ＮＰ」になるという構文規則が適用されて状態スタック６３に示すように名詞句の「ＮＰ」に変換される。ここで、状態スタック６３の名詞句の「ＮＰ」は、「きれいな」が「女の子を」に係る構造の「きれいな黒い髪の女の子を」を表す。さらに、状態スタック６４に示すように文章を表す「Ｓ」に変換されて、「きれいな」が「女の子を」に係るような構造の音声認識結果のみが出力される。以上のようにポーズを表示する「△」の前後に位置する「きれいな」を表す形容詞句の「ａｄｊ」と「黒い髪の女の子を」を表す名詞句の「ＮＰ」との間の構文規則の適用を最後のポーズ単位の末端まで到達してから実行するので、「きれいな」が「女の子を」に係る構造の音声認識結果のみを出力することができる。以上のように第１の実施形態では、複数の音声区間からなる入力された発声音声の１つの文の各音声区間の音声認識を実行した後、区間を限定せず、入力された発声音声の１つの文に対して異なる音声区間に属する語、句又は節の間にＬＲ[illegible]ブルメモリ１３内の構文規則を適用して

異なる音声区間に属する語、句又は節の間の係り受け関係を決定している。

【００３２】図４は、図２の例文において第２の係り受け関係を有する場合の連続音声認識装置８１の音声認識動作をスタック形式で示す図である。以下に第２の係り受け関係を有する場合の音声認識動作を図４を参照して説明する。まず、図４の状態スタック１５１に示すように、ＬＲパーザ５で「きれいな」という発声音声の系列が認識されて文字として積まれる。次に、状態スタック１５１における「きれいな」という文字は音声認識用辞書に載っているので、状態スタック１５２に示すように形容詞句を表す「ａｄｊ」という文字に変換される。次に、ＬＲパーザ５で「黒い」という発声音声の系列が認識されて状態スタック１５３に示すように「きれいな」を表す形容詞句の「ａｄｊ」の上に文字として積まれ、状態スタック１５３における「黒い」は音声認識辞書に載っているので状態スタック１５４に示すように形容詞句を表す「ａｄｊ」という文字に変換される。

【００３３】次に、ＬＲパーザ５で「髪の」という発声音声の系列が認識されて状態スタック１５５に示すように「黒い」を表す形容詞句の「ａｄｊ」の上に文字として積まれ、次に「髪の」の認識処理の直後でポーズが無音区間等検出部３０によって検出されて、検出信号が当該検出部３０からＬＲパーザ５に入力されてポーズを表示する「△」として「髪の」の文字の上に積まれる。そして、状態スタック１５５における「髪の」という文字は音声認識辞書に載っているので状態スタック１５６に示すように名詞句を表す「ＮＰ」という文字に変換される。

【００３４】さらに、状態スタック１５６において、形容詞句の「ａｄｊ」と名詞句の「ＮＰ」とは名詞句の「ＮＰ」になるという構文規則が適用されて、「黒い」が変換された形容詞句の「ａｄｊ」と「髪の」が変換された名詞句の「ＮＰ」とは状態スタック１５７に示すように名詞句の「ＮＰ」に変換される。すなわち、状態スタック１５７における名詞句の「ＮＰ」は「黒い髪の」を表す。さらに、状態スタック１５７において、形容詞句の「ａｄｊ」と名詞句の「ＮＰ」とは名詞句の「ＮＰ」になるという構文規則が適用されて、「きれいな」が変換された形容詞句の「ａｄｊ」と「黒い髪の」を表す名詞句の「ＮＰ」とは状態スタック１５８に示すように名詞句の「ＮＰ」に変換される。これによって、「きれいな」が「髪の」にかかる構造として認識される。

【００３５】次に、「女の子を」という発声音声の系列が認識されて状態スタック１５９に示すようにポーズを表示する「△」の上に文字として積まれ、状態スタック１５９における「女の子を」の文字は音声認識辞書に載っているので状態スタック１６０に示すように名詞句を表す「ＮＰ」という文字に変換される。ここで、状態スタック１６０において「きれいな黒い髪の」を表す名詞

句の「ＮＰ」と「女の子を」が変換された名詞句の「ＮＰ」とには、間にポーズを表示する「△」が積まれているので構文規則は適用されない。そして、「見た」という発声音声の系列が認識されて状態スタック１６１に示すように「女の子を」が変換された名詞句の「ＮＰ」の上に文字として積まれ、状態スタック１６１における「見た」は音声認識用辞書に載っているので状態スタック１６２に示すように動詞句を表す「ＶＰ」に変換される。

【００３６】そして、ＬＲパーザ５で最後のポーズ単位の末端まで到達していると判断されて、状態スタック１６２におけるポーズを表示する「△」の前後に位置する「きれいな黒い髪の」を表す名詞句の「ＮＰ」と「女の子を」を表す名詞句の「ＮＰ」とにＬＲテーブルメモリ１３内の構文規則が適用されて、状態スタック１６３に示すように名詞句の「ＮＰ」に変換されて、さらに状態スタック１６４に示すように文章を表す「Ｓ」の文字に変換されて、「きれいな」が「黒い髪の」に係るような構造の音声認識結果のみが出力される。

【００３７】以上の第１の実施形態の連続音声認識装置８１は、無音区間等を検出して検出信号を出力する無音区間等検出部３０を備え、ＬＲパーザ５は、無音区間等検出部３０から入力された検出信号で示された区切り単位の音声区間のデータを読み込んで、当該音声区間に対してＨＭＭ－ＬＲ法を用いた区間制限付きＨＭＭ－ＬＲ処理を実行し、最後の区切り単位の末端まで到達すると入力された発声音声の１つの文に対して区間制限無しＨＭＭ－ＬＲ処理を実行することにより音声認識結果データを出力する。これによって、各区切り単位の音声区間内における係り受け関係を決定した後、異なる区切り単位の音声区間に属する語、句又は節の間の係り受け関係を決定できるので、統語解析における係り受け関係の曖昧性を解消することができる。

【００３８】<第２の実施形態>図７は、本発明に係る第２の実施形態である連続音声認識装置８２のブロック図である。図７の第２の実施形態の連続音声認識装置８２は、図１の第１の実施形態の連続音声認識装置８１の隠れマルコフ網メモリ１１に代えて隠れマルコフ網メモリ１１ａを備え、かつ無音区間等検出部３０を除いて構成される。第２の実施形態の連続音声認識装置８２においては、ポーズや冗長語並びに韻律的な情報等を手がかりとする区切りなどの無音区間等をＨＭＭでモデル化したモデルが隠れマルコフ網メモリ１１ａに格納され、当該モデルを用いて無音区間等の検出を音素照合部４で行っている。

【００３９】図８は、図７の連続音声認識装置８２において実行される音声認識処理を示すフローチャートである。以下、図８を参照して第２の実施形態の連続音声認識装置８２の音声認識処理について説明する。まず、ス

にＬＲパーザ５の初期化を実行する。具体的には、状態スタック０のセルを１個作成する。そして、ステップＳ１１において、例えば、特徴パラメータの処理単位である音声フレーム（例えば２０ミリ秒）毎に音声データの読み込みを行い、ステップＳ１２において区間制限付きＨＭＭ－ＬＲ処理を実行する。次にステップＳ１３において無音区間等を検出したか否かが判断され、無音区間等を検出していない場合はステップＳ１１に進みステップＳ１１，Ｓ１２の処理が繰り返され、無音区間等を検出した場合はステップＳ１４に進む。

【００４０】ステップＳ１４において、すべての音声区間の音声認識処理が終了したか否かが判断され、すべての音声区間の処理が終了していないときは（ステップＳ１４においてＮＯ）ステップＳ１１に進み、ステップＳ１１，Ｓ１２，Ｓ１３の処理を繰り返し、すべての音声区間の処理が終了したと判断されると（ステップＳ１４においてＹＥＳ）ステップＳ１５に進み、入力された発声音声の１つの文に対して区間制限無しＨＭＭ－ＬＲ処理を実行して音声認識処理を終了する。

【００４１】以上の第２の実施形態の連続音声認識装置８２は、無音区間等の検出を隠れマルコフ網メモリ１１ａに格納されたＨＭＭでモデル化した無音区間等のモデルを使用して音素照合部４で行い、ＬＲパーザ５は、音声データを読み込んで、１つの音声区間に対してＨＭＭ－ＬＲ法を用いた区間制限付きＨＭＭ－ＬＲ処理を実行し、各音声区間についての処理が終了すると入力された発声音声の１つの文に対して区間制限無しＨＭＭ－ＬＲ処理を実行することにより音声認識結果データを出力する。これによって、各区切り単位の音声区間内における係り受け関係を決定した後、異なる音声区間に属する語、句又は節の間の係り受け関係を決定できるので、統語解析における係り受け関係の曖昧性を解消することができる。

【００４２】以上の第１と第２の実施形態においては、入力された発声音声の１つの文に対して区間制限無しＨＭＭ－ＬＲ処理を実行することにより音声認識結果データを出力するようにした。しかしながら、本発明はこれに限らず、入力された発声音声の１つの句又は節等の１つのシーケンスの発声音声に対して区間制限無しＨＭＭ－ＬＲ処理を実行するようにしてもよいし、連続音声認識装置のスイッチがオンされてからオフされるまでの間に入力される発声音声に対して区間制限無しＨＭＭ－ＬＲ処理を実行するようにしてもよい。以上のように構成しても第１と第２の実施形態と同様に動作し同様の効果を有する。

【００４３】以上の第１と第２の実施形態においては、ＨＭＭ－ＬＲ法を用いた音声認識装置について述べているが、本発明はこれに限らず、ニューラルネットワークを用いた音声認識装置など他の種類の音声認識装置に適

13

【００４４】

【発明の効果】本発明に係る請求項１記載の音声認識装置は、入力された発声音声に基づいてポーズと冗長語と句又は節の境界とのうちの少なくとも１つを検出して検出信号を出力する検出手段を備え、上記音声認識手段は、上記検出信号に基づいて統語解析における係り受け関係を決定して上記発声音声の音声認識をしている。これによって、統語解析における係り受け関係の曖昧性を解消できる。

【００４５】また、請求項２記載の音声認識装置は、請求項１記載の音声認識装置において、上記音声認識手段は、上記ポーズと冗長語と句又は節の境界とのうちの少なくとも１つによって分割された複数の音声区間からなる入力された発声音声の各音声区間について音声認識処理をした後、異なる音声区間に属する語、句又は節の間の係り受け関係を決定して、上記入力された発声音声の音声認識をしている。これによって、統語解析における係り受け関係の曖昧性を解消できる。

【００４６】さらに、請求項３記載の音声認識装置は、請求項１又は２記載の音声認識装置において、上記検出手段は、上記発声音声のパワーが、所定の時間の範囲だけ、所定のしきい値以下である第１の条件と、上記発声音声のゼロクロスの数が、所定の時間の間において、所定のしきい値以上である第２の条件とのうち少なくとも１つの条件が満足することを検出することにより上記ポーズを検出している。これによって、上記ポーズに基づいて統語解析における係り受け関係を決定でき、統語解析における係り受け関係の曖昧性を解消できる。

【００４７】またさらに、請求項４記載の音声認識装置は、請求項１又は２記載の音声認識装置において、上記検出手段は、上記ポーズと冗長語と句又は節の境界とのうちの少なくとも１つを、それぞれの予め決められた言語モデルに一致するか否かを判断することにより検出している。これによって、音声認識過程で上記ポーズと冗長語と句又は節の境界とのうちの少なくとも１つを検出

14

でき、統語解析における係り受け関係の曖昧性を解消できる。

【図面の簡単な説明】

【図１】　本発明に係る第１の実施形態である連続音声認識装置のブロック図である。

【図２】　図１の連続音声認識装置８１の音声認識動作を説明するために用いた第１と第２の２つの係り受け関係を有する一例文を示す図である。

【図３】　図１の連続音声認識装置８１の音声認識動作の一例をスタック形式で示す図である。

【図４】　図１の連続音声認識装置８１の音声認識動作の図３とは異なる例をスタック形式で示す図である。

【図５】　図２の例文の音声区間、ポーズ（無音区間等）及びポーズ単位（区切り単位）を示す図である。

【図６】　図１の連続音声認識装置８１のＬＲパーザ５によって実行される音声認識処理を示すフローチャートである。

【図７】　本発明に係る第２の実施形態である連続音声認識装置８２のブロック図である。

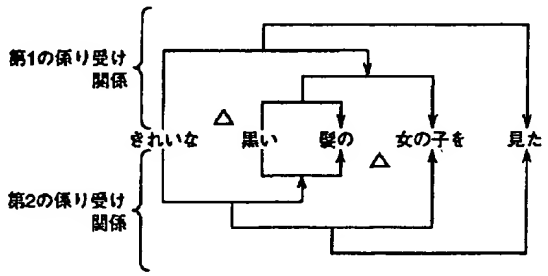
【図８】　図７の連続音声認識装置８２のＬＲパーザ５によって実行される音声認識処理を示すフローチャートである。

【図９】　従来例の連続音声認識装置の音声認識動作をスタック形式で示す図である。

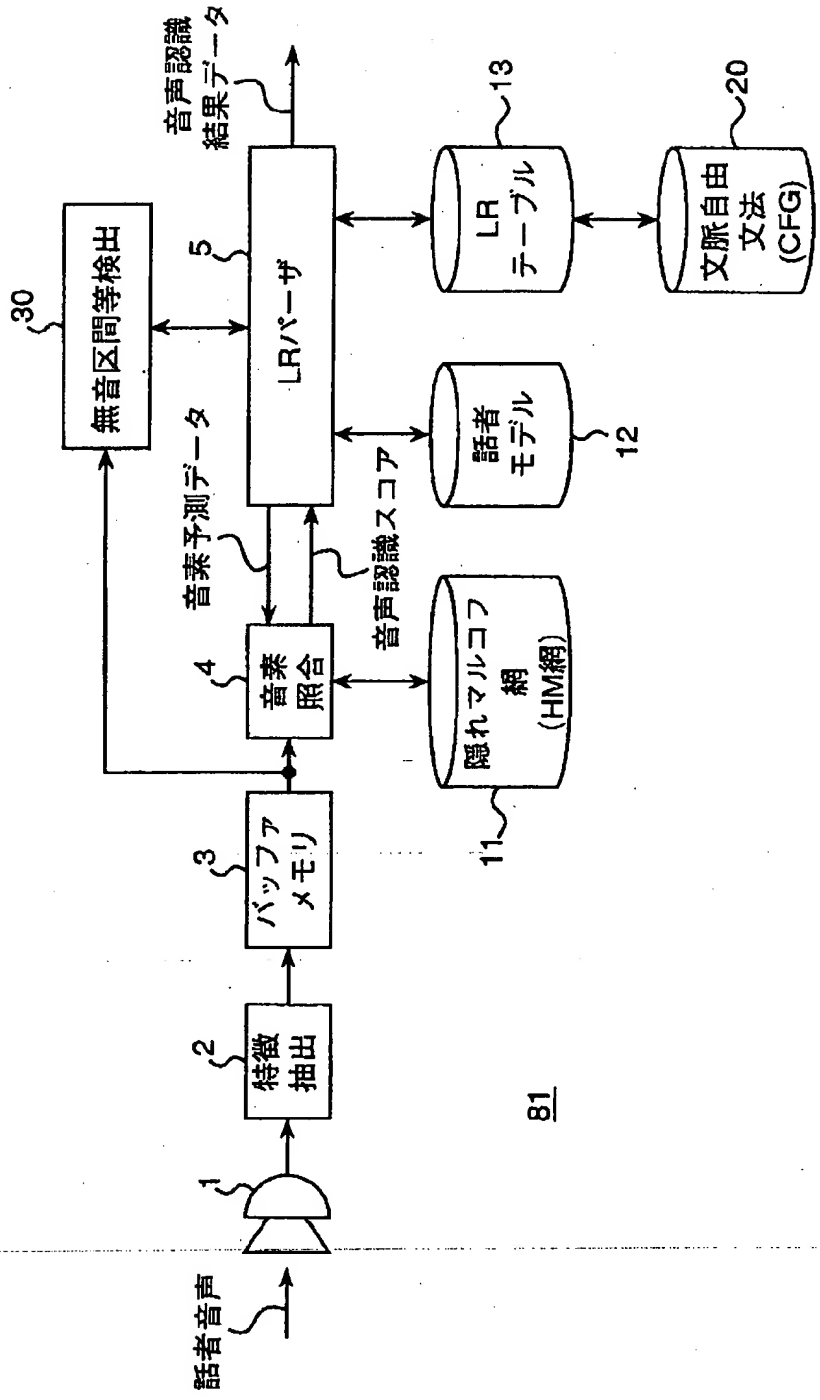
【符号の説明】

１…マイクロホン、
２…特徴抽出部、
３…バッファメモリ、
４…音素照合部、
５…ＬＲパーザ、
１１，１１ａ…隠れマルコフ網メモリ、
１２…話者モデルメモリ、
１３…ＬＲテーブルメモリ、
２０…文脈自由文法データベースメモリ、
３０…無音区間等検出部。

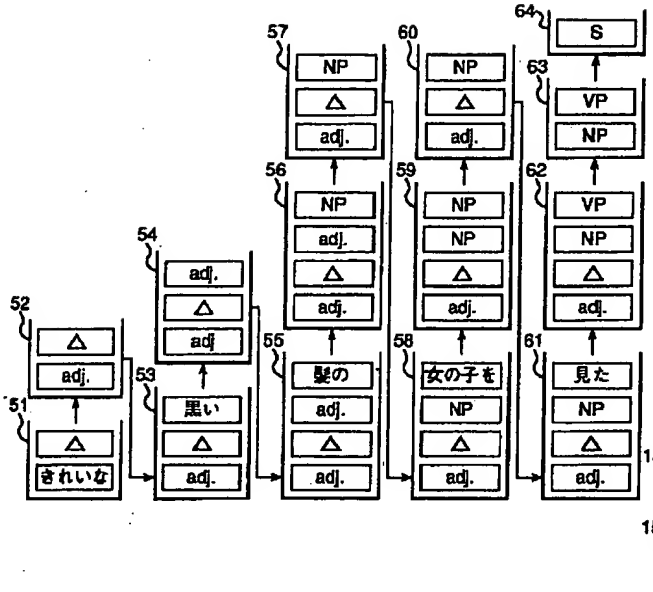
【図２】



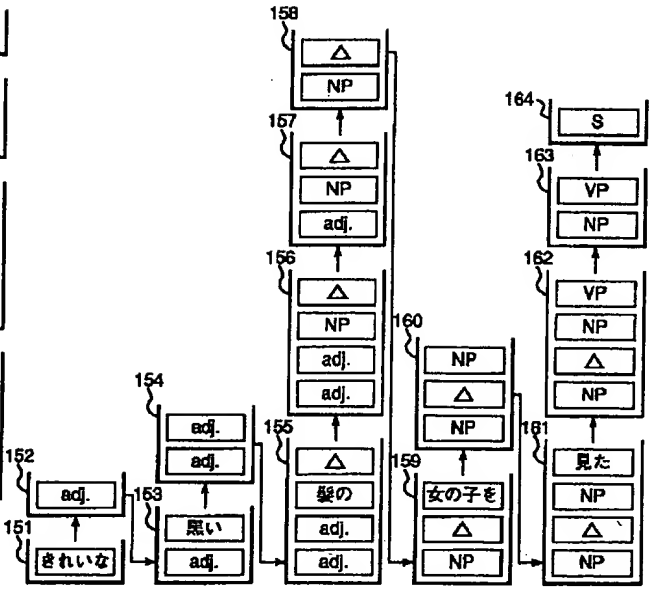
【図１】


話者音声
1
特徴抽出
2
バッファメモリ
3
音素照合
4
音素予測データ
音声認識スコア
隠れマルコフ網(HM網)
11
話者モデル
12
LRテーブル
13
文脈自由文法(CFG)
20
LRパーザ
5
無音区間等検出
30
音声認識結果データ
81

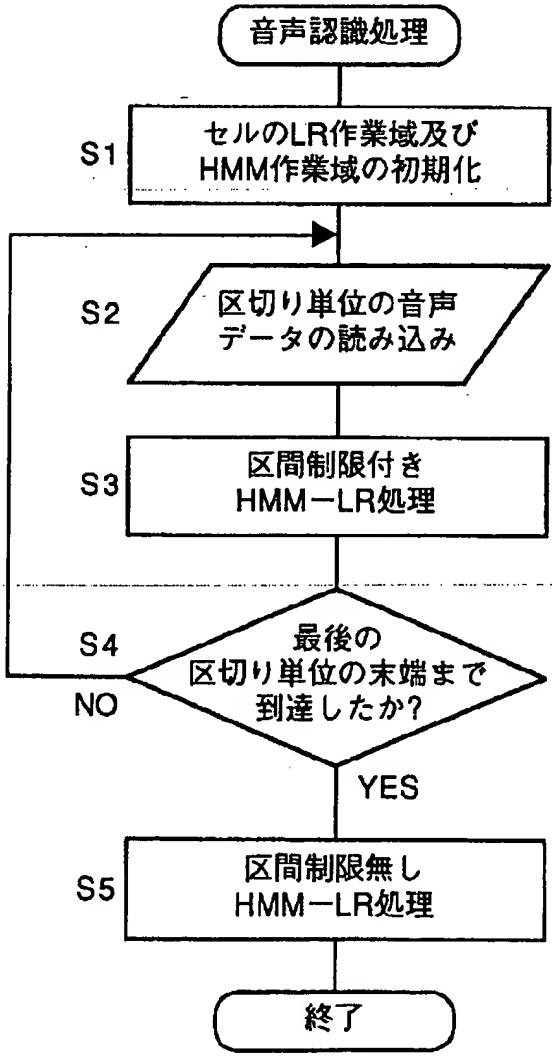
【図3】


51
52
53
54
55
56
57
58
59
60
61
62
63
64
きれいな
adj.
△
黒い
髪の
女の子を
見た
NP
VP
S

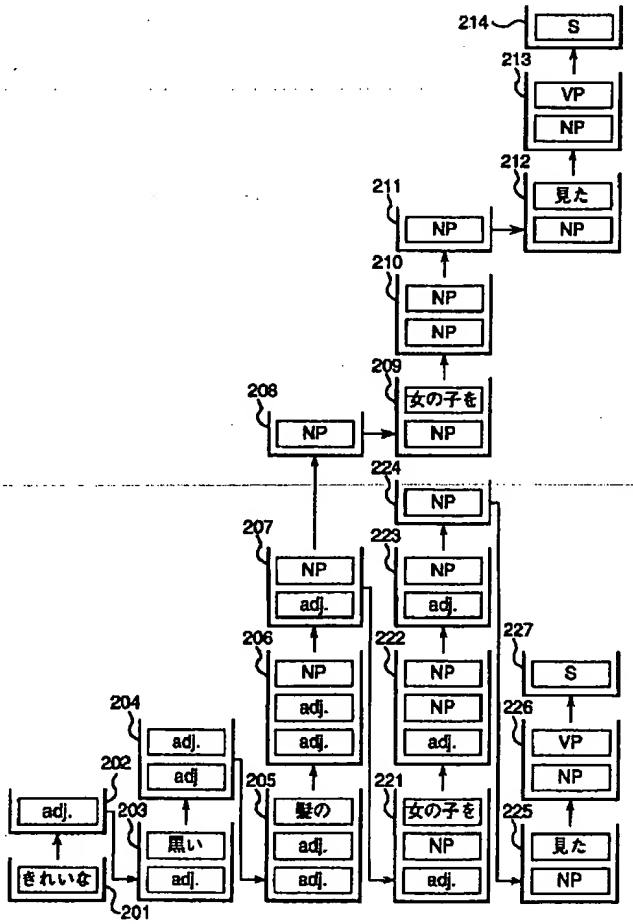
【図4】


151
152
153
154
155
156
157
158
159
160
161
162
163
164
きれいな
黒い
髪の
女の子を
見た
adj.
△
NP
VP
S

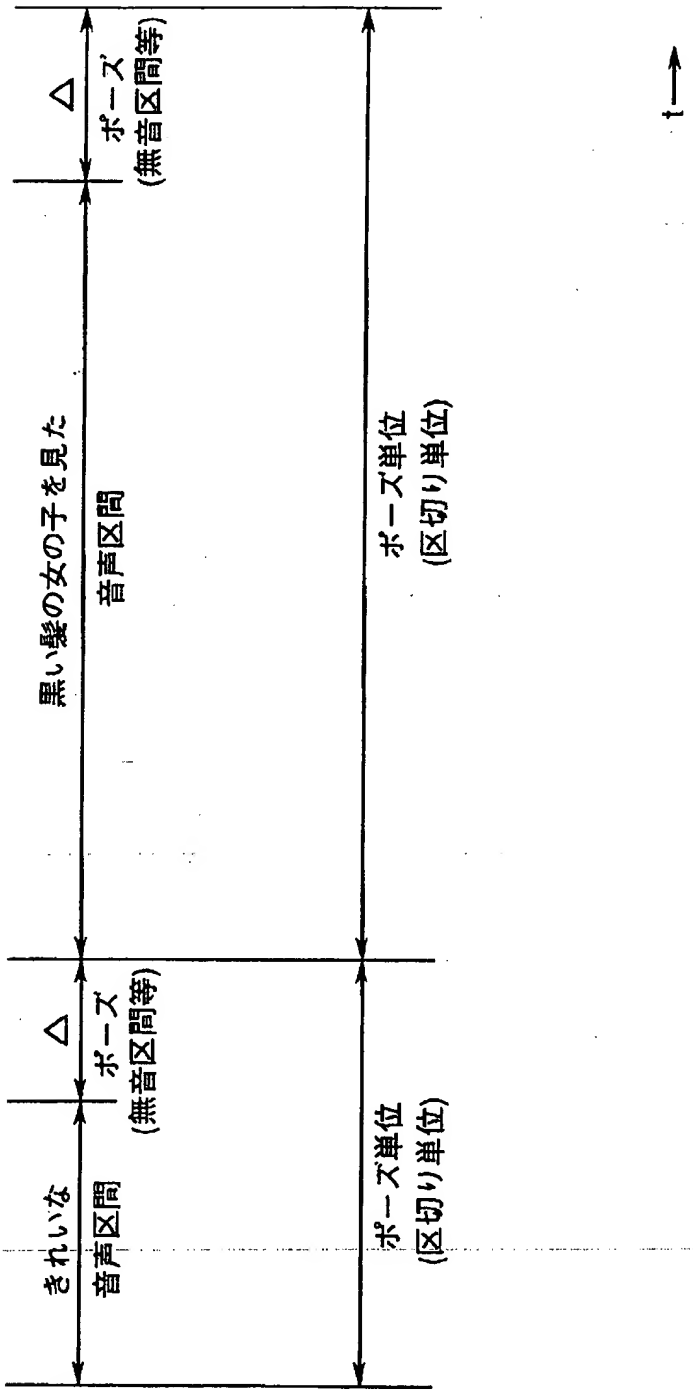
【図6】


音声認識処理
S1
セルのLR作業域及び
HMM作業域の初期化
S2
区切り単位の音声
データの読み込み
S3
区間制限付き
HMM－LR処理
S4
最後の
区切り単位の末端まで
到達したか?
NO
YES
S5
区間制限無し
HMM－LR処理
終了

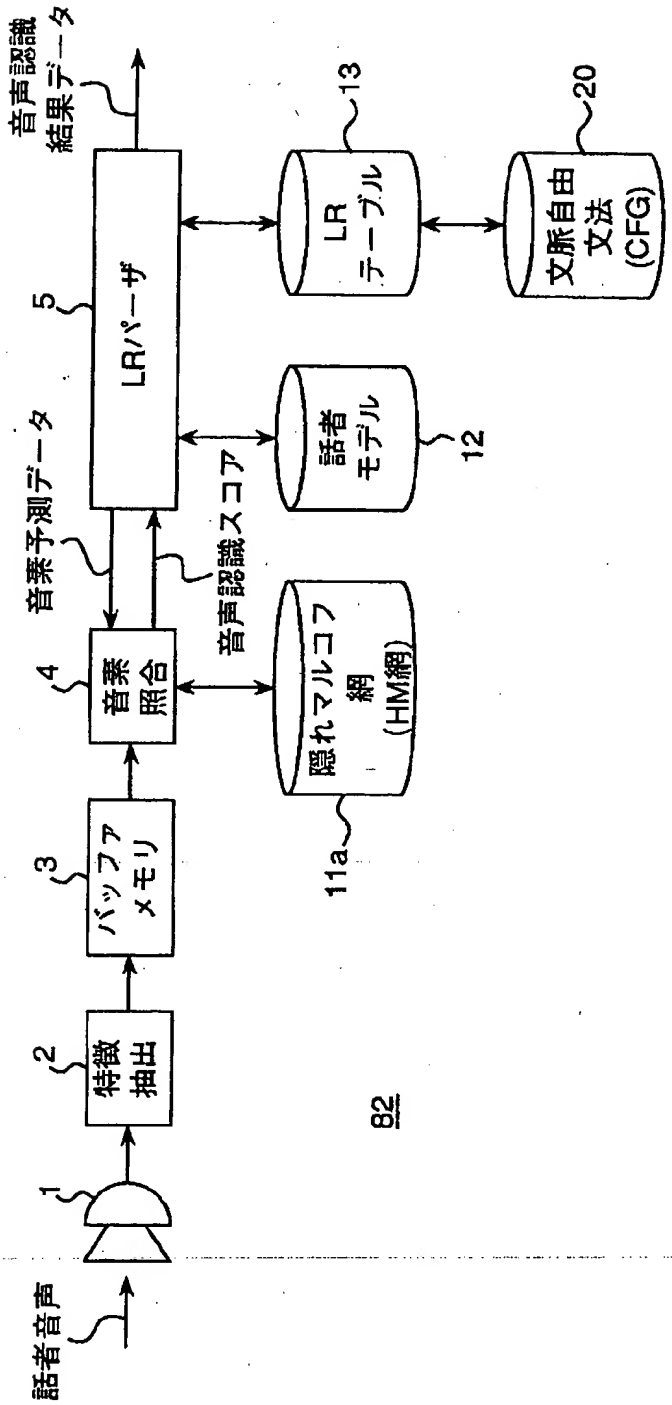
【図9】


201
202
203
204
205
206
207
208
209
210
211
212
213
214
221
222
223
224
225
226
227
きれいな
黒い
髪の
女の子を
見た
adj.
△
NP
VP
S

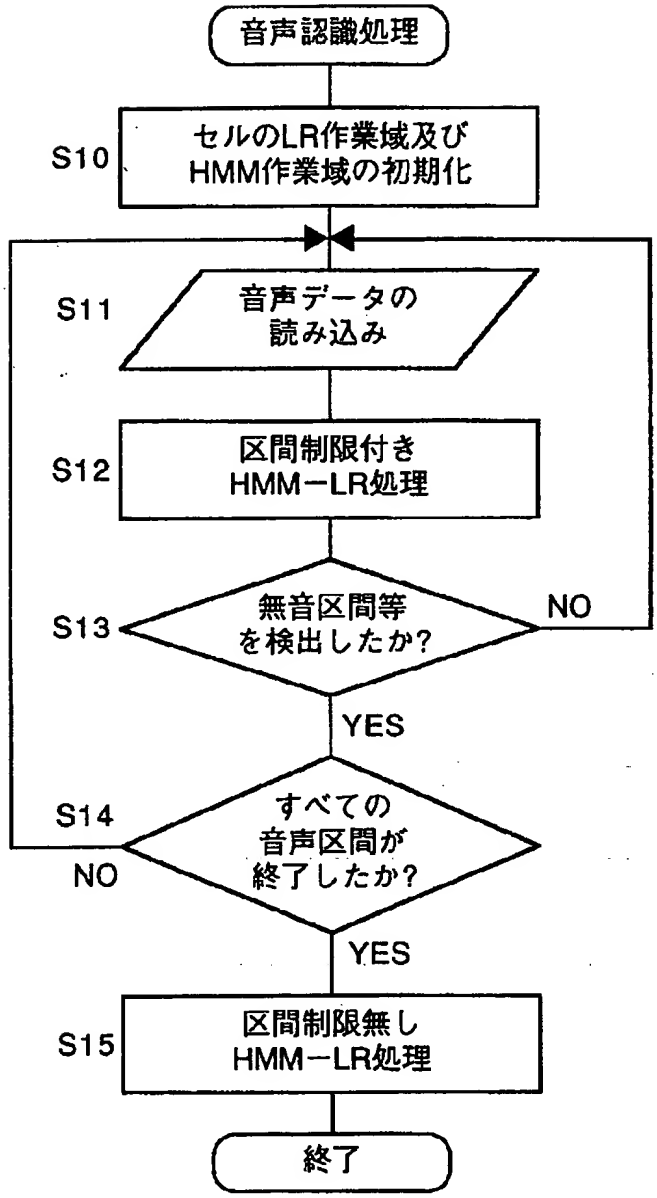
【図５】


△
ポーズ
(無音区間等)
黒い髪の女の子を見た
音声区間
ポーズ単位
(区切り単位)
△
ポーズ
(無音区間等)
きれいな
音声区間
ポーズ単位
(区切り単位)
t

【図７】


話者音声
1
2
特徴抽出
3
バッファメモリ
4
音素照合
音素予測データ
音声認識スコア
5
LRパーザ
音声認識結果データ
隠れマルコフ網(HM網)
11a
話者モデル
12
LRテーブル
13
文脈自由文法(CFG)
20
82

【図８】


音声認識処理
S10
セルのLR作業域及び
HMM作業域の初期化
S11
音声データの
読み込み
S12
区間制限付き
HMM－LR処理
S13
無音区間等
を検出したか?
NO
YES
S14
すべての
音声区間が
終了したか?
NO
YES
S15
区間制限無し
HMM－LR処理
終了